



**University of  
Zurich**<sup>UZH</sup>

**Zurich Open Repository and  
Archive**

University of Zurich  
University Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2018

---

## **Can self-defeating humor make you happy? Cognitive interviews reveal the adaptive side of the self-defeating humor style**

Heintz, Sonja ; Ruch, Willibald

**Abstract:** The present set of studies employs two cognitive interviewing techniques (thinking aloud and online cognitive probing) of the scale assessing the self-defeating humor style, aiming at delineating the role that self-defeating humor plays in self-esteem and emotions. The self-defeating humor style comprises humor to enhance one's relationships with others at the expense of oneself, and has often been related to lower well-being. The analyses are based on 392 item responses of a typical sample (Study 1) and 104 item responses of high scorers on the self-defeating scale (Study 2). Content analyses revealed that higher scores on the self-defeating scale went along with humor (Study 1), with higher state self-esteem, with an improvement of one's interpersonal relationships, and with more facial displays of positive emotions (Study 2). Additionally, the more humor was entailed in the item responses, the higher the state self-esteem and the improvement of relationships was and the more positive emotion words were employed. Thus, the humor entailed in the self-defeating humor style seemed rather beneficial both for oneself and others. These findings call for a reevaluation of past findings with this humor style and provide opportunities for future research and applications of humor interventions to improve well-being.

DOI: <https://doi.org/10.1515/humor-2017-0089>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-151378>

Journal Article

Published Version

Originally published at:

Heintz, Sonja; Ruch, Willibald (2018). Can self-defeating humor make you happy? Cognitive interviews reveal the adaptive side of the self-defeating humor style. *HUMOR: International Journal of Humor Research*, 31(3):451-472.

DOI: <https://doi.org/10.1515/humor-2017-0089>

Sonja Heintz\* and Willibald Ruch

# Can self-defeating humor make you happy? Cognitive interviews reveal the adaptive side of the self-defeating humor style

<https://doi.org/10.1515/humor-2017-0089>

**Abstract:** The present set of studies employs two cognitive interviewing techniques (thinking aloud and online cognitive probing) of the scale assessing the self-defeating humor style, aiming at delineating the role that self-defeating humor plays in self-esteem and emotions. The self-defeating humor style comprises humor to enhance one's relationships with others at the expense of oneself, and has often been related to lower well-being. The analyses are based on 392 item responses of a typical sample (Study 1) and 104 item responses of high scorers on the self-defeating scale (Study 2). Content analyses revealed that higher scores on the self-defeating scale went along with humor (Study 1), with higher state self-esteem, with an improvement of one's interpersonal relationships, and with more facial displays of positive emotions (Study 2). Additionally, the more humor was entailed in the item responses, the higher the state self-esteem and the improvement of relationships was and the more positive emotion words were employed. Thus, the humor entailed in the self-defeating humor style seemed rather beneficial both for oneself and others. These findings call for a reevaluation of past findings with this humor style and provide opportunities for future research and applications of humor interventions to improve well-being.

**Keywords:** self-defeating humor style, Humor Styles Questionnaire, self-esteem, emotions, cognitive interviews, self-directed humor

## 1 Introduction

Making fun of oneself has often been conceptualized as a positive trait, labeled laughing at oneself (see McGhee 2010). This concept entails that no human is

---

**\*Corresponding author: Sonja Heintz**, Department of Psychology, Personality and Assessment, University of Zurich, Binzmuehlestrasse 14, Box 7, Zürich CH-8050, Switzerland,  
E-mail: [s.heintz@psychologie.uzh](mailto:s.heintz@psychologie.uzh)

**Willibald Ruch**, Department of Psychology, Personality and Assessment, University of Zurich, Binzmuehlestrasse 14, Box 7, Zürich CH-8050, Switzerland

perfect, that we all make mistakes once in while, and that we do not need to react negatively to them. Consequently, laughing at oneself should reflect better psychological well-being. By contrast, it was proposed and empirically supported in recent years that habitually making fun of oneself excessively (i.e., a self-defeating humor style) can have detrimental effects on one's psychological well-being (see Kuiper 2014; Martin et al. 2003). The present set of studies employs cognitive interviewing techniques that tap directly into the cognitive processes underlying the responses to the items of the self-defeating humor style. This allows delineating the relationship of the self-defeating humor style and the humor entailed in it with psychological well-being (state self-esteem and emotions), enabling a more direct evaluation of how adaptive or maladaptive self-defeating humor actually is.

## 1.1 Self-defeating humor style

Martin et al. (2003: 48) proposed the self-defeating humor style, defined as the “use of humor to enhance relationships at the expense of self.” This humor style is characterized by excessive self-disparaging humor, by making fun of oneself to be approved by others, and by using humor to suppress one's feelings. Martin et al. (2003: 52–54) further outlined that “Self-defeating humor is seen as potentially detrimental to well-being when used excessively, since it involves denigration of the self and repression of one's own emotional needs” and that “there is an element of emotional neediness, avoidance, and low self-esteem underlying their use of humor.” Thus, this humor style is negatively connoted, both in terms of self-esteem and emotions.

Martin et al. (2003) also presented a self-report measure (the Humor Styles Questionnaire, HSQ) to assess the self-defeating humor style, along with three other humor styles (affiliative, self-enhancing, and aggressive). The self-defeating (SD) scale entails eight items (e.g., “I often try to make people like or accept me more by saying something funny about my own weaknesses, blunders, or faults”) that are answered on a 7-point Likert-type scale (from “totally disagree” to “totally agree”). In line with the conceptualization of the self-defeating humor style, empirical findings with the SD scale supported its relationships with lower well-being. For example, the SD scale was found to correlate negatively with psychological well-being (lower trait positive affect and global self-esteem and more trait negative affect; e.g., Kuiper 2014; Maiolino and Kuiper 2014; Martin et al. 2003; Rnic et al. 2016; Ruch and Heintz 2013; Ruch and Heintz 2017) and with maladaptive outcomes for others, such as hostility and antagonism (e.g., Martin et al. 2003; Zeigler-Hill et al. 2016).

Several observations can be made regarding these findings. First, average SD scores usually ranged between 24–30 ( $SD = 8-10$ ), with 32 being the scale midpoint (i.e., “neither agree nor disagree”). Thus, most participants tended to on average disagree with the SD items. In other words, the findings mostly apply to low and medium scorers in the SD scale, while the high scorers (i.e., those who, on average, tend to agree with the items) have not been explored thus far. Second, the studies were mostly correlational, cross-sectional, and relied on self-reports. Thus, the causal relationships between the SD scale and well-being and the generalizability across different methods still needs to be tested.

## 1.2 Self-defeating humor

Importantly, the findings with the SD scale were often attributed to the self-defeating humor entailed in it. For example, Rnic et al. (2016: 358) concluded that “The use of Self-Defeating humor may reinforce the individual’s negative self-concept (thereby increasing negative affect), especially when others appear to agree with the individual’s humorous actions or statements, or to react to their use of humor in a rejecting manner.” Consequently, some authors advised to identify and decrease the use of self-defeating humor as a maladaptive behavior (e.g., Maiolino and Kuiper 2014; Rnic et al. 2016). These conclusions rest on the assumption that the SD scale actually captures self-defeating humor and that engaging in this self-defeating humor causes or adds to the maladaptive outcomes and reduces well-being, either directly or by negative social reinforcements and consequences.

Recent research investigated these assumptions, casting doubt on the maladaptive nature of self-defeating humor. Specifically, Ford et al. (2017) induced anxiety using an imagery task and then exposed participants to a self-defeating humor manipulation (using written instructions, jokes, and cartoons). A measure of state anxiety was administered after the manipulation. They found that engaging in self-defeating humor did not increase anxiety in comparison to a no-humor control group. Using a daily diary paradigm, Heintz (2017) investigated everyday humor behaviors and found that the humor behaviors entailed in the SD scale were unrelated to positive and negative affect, and that the SD scale did not significantly correlate with its daily measured humor behaviors.

Another line of research experimentally manipulated the HSQ items, so they contained only the humor content (Ruch and Heintz 2013; 2017). For example, the item “I often go overboard in putting myself down when I am making jokes or trying to be funny” was turned into a humor-only version (“I often make jokes about myself or make fun of myself”). Ruch and Heintz (2013) found that the humor-only version of the SD scale was uncorrelated with self-esteem and affect,

but positively correlated with adaptive forms of humor (e.g., coping humor and laughing at oneself). One possibility the authors discussed in explaining their findings is that the SD scale contains a positive laughing at oneself and additionally a negative component (like neuroticism or low self-esteem). Specifically, the latter might be entailed in the items by terms such as “going overboard” or “more than one should.” Thus, the negative connotation of the SD scale might not be due to actually showing excessive self-disparaging humor (i.e., too frequently, too intensely, or both), but it might be due to a lower self-esteem or a proneness to negative affect that generalizes to everything the person does. The present studies extend these findings by directly investigating the interplay of humor, self-esteem, and emotions in the SD items.

### 1.3 Self-esteem and emotion expressions

Three well-being related outcomes are relevant for the present studies: Self-esteem, emotion words, and facial displays of emotion. First, self-esteem can be regarded as a global self-evaluation, ranging from positive to negative attitudes toward oneself (Rosenberg et al. 1995). Self-esteem can be conceptualized as a state (how positive vs. negative a person’s self-evaluation is at the moment) or as a trait (how positive vs. negative it is in general). Second, people differ in how many words indicating positive or negative emotions they use. For example, the use of negative emotion words was related to lower levels of self-esteem (Bosson et al. 2000). Third, universal facial displays of a variety of emotions have been identified, and their relationship with subjective emotional experience was supported in several studies (for an overview, see Matsumoto et al. 2010). The present studies are the first to investigate how the SD scale and self-defeating humor directly relate to state self-esteem, to the use of emotion words, and to facial displays of emotion. This extends previous self-report findings to content ratings and to behavior-based measures of emotion.

### 1.4 Cognitive interviewing

Cognitive interviewing is defined as “the administration of draft survey questions while collecting additional verbal information about the survey responses, which is used to evaluate the quality of the response or to help determine whether the question is generating the information that its author intends” (Beatty and Willis 2007: 287). Two basic techniques are differentiated, namely thinking aloud and probing. They mainly differ in their timing: Thinking aloud requires participants to

continuously report their thoughts while answering items, while probing consists of asking specific questions after the response to an item was given (Beatty and Willis 2007). Recently, online probing was introduced as a fruitful addition to classic probing to elicit item explanations in larger samples (e.g., Behr et al. 2012). While cognitive interviewing is usually employed during item construction, this technique can also reveal information on the cognitive processes underlying existing scales. The present studies employ two cognitive interviewing techniques, online probing and thinking aloud, of the SD scale.

## 1.5 Aims of the Present Studies

First, cognitive interviewing of the SD items yields information on how they are processed and interpreted. Study 1 employs online cognitive probing in a standard sample (in which participants explain why they chose a specific answer option to each SD item). Study 2 employs thinking aloud with high scorers on the SD scale to cover the full spectrum of scores (and not only low to medium scorers as in standard samples). Second, these responses to each of the SD items are rated regarding the amount to which they entail self-directed humor (from none to a lot) and regarding the state self-esteem expressed in the response (positive, negative, or neutral self-evaluation). Third, the item responses are analyzed to yield the relative frequencies with which positive and negative emotion words are used. Fourth, participants in Study 2 are filmed and their facial displays of emotion during the responses are analyzed.

Two sets of hypotheses are put up regarding the scores and the humor entailed in the SD items. In line with Martin et al.'s (2003) notion of the self-defeating humor style, responses of higher scores in the SD items should contain more self-directed humor and should be more maladaptive (i.e., entail a negative self-evaluation, negative emotion words, and be accompanied by facial displays of negative emotion). For example, participants who agree with the item "I often go overboard in putting myself down when I am making jokes or trying to be funny" should report to often go overboard in putting themselves down (negative self-evaluation, or low state self-esteem) and to often make jokes or try to be funny (humor). Also, their responses should be accompanied by negative emotions, because they might feel uncomfortable and sad when thinking about their experiences that made them chose the answer. These emotions should be behaviorally expressed by using more negative emotion words and by showing more facial displays of negative emotion in comparison to people who disagree with the item (as they can disagree because they do not go overboard in putting themselves down and/or not often make jokes or try to be funny).

Based on the recent studies that suggested that the humor entailed in the SD scale is neutral or adaptive, it is expected that the amount of self-directed humor entailed in the responses to the SD items (as determined by the content ratings) should be either uncorrelated or negatively correlated with the frequencies of facial displays of negative emotion and with negative emotion words, and either be uncorrelated or positively correlated with state self-esteem, facial displays of positive emotion, and positive emotion words.

## 1.6 Preliminary study

To select participants for Studies 1 and 2, two online samples were collected, in which participants completed several questionnaires on humor and psychological well-being (Sample 1:  $N=218$ , Sample 2:  $N=502$ ). These data served as convergent validation for the SD scale employed in Studies 1 and 2, and as a selection criterion for high scorers on the SD scale for Study 2 (i.e., participants with scores  $\geq 33$  on a scale from 8–56).

Participants in both samples completed the HSQ (Martin et al. 2003), which assesses four humor styles (affiliative, self-enhancing, aggressive, and self-defeating) using a seven-point Likert-type scale. Only the eight-item SD scale is relevant for the present studies, for which the internal consistencies (Cronbach's alpha) were sufficient for the sample of Study 1 (0.80) and Study 2 (0.61). Participants also completed trait versions of the relevant psychological well-being variables (i.e., self-esteem and emotions): The Rosenberg Self-Esteem Scale (Rosenberg 1965), which measures general self-esteem with 10 items (employing a four-point Likert-type scale), and the Scale of Positive and Negative Experience (Diener et al. 2010), which employs six items each to measure the frequency of positive and negative feelings across the past four weeks (on a five-point Likert-type scale). Cronbach's alpha was sufficient for self-esteem (Study 1: 0.93, Study 2: 0.94), for positive feelings (Study 1: 0.93, Study 2: 0.97), and for negative feelings (Study 1: 0.82, Study 2: 0.90).

## 2 Study 1: Online Cognitive Probing of the Self-Defeating Humor Scale

Study 1 employs online cognitive probing to determine why participants select an answer option in the SD items. Specifically, the SD item scores are related to content ratings (humor and self-esteem) and frequencies of emotion words

entailed in the responses to the probing questions. Regarding the content ratings, it is expected that the SD item scores correlate positively with the amount of self-directed humor (Hypothesis 1.1) and negatively with state self-esteem (Hypothesis 1.2). Regarding the frequencies of emotion words, it is expected that the SD item scores correlate positively with negative emotion words (Hypothesis 1.3). In terms of the humor entailed in the responses, it is expected that higher scores in the self-directed humor content rating are either uncorrelated or correlate positively with state self-esteem (Hypothesis 1.4) and the use of positive emotion words (Hypothesis 1.5), and that they are either uncorrelated or negatively correlated with the use of negative emotion words (Hypothesis 1.6). Hypotheses 1.4–1.6 are tested across both all item responses and across the items that were agreed to, as the affirmed responses are the ones that should represent the self-defeating humor style most strongly.

## 2.1 Method

### 2.1.1 Sample

Of the 81 people who agreed to participate, 70 completed the study and provided meaningful answers in the cognitive probing (86.4%). In a consistency check, 21 of them were excluded because their SD scores from the preliminary study and Study 1 differed eight points or more from one another,<sup>1</sup> resulting in a final sample of 49 participants (SD score ranging from 9–46), who provided a total of 392 responses. Eight men and 41 women participated. Their average age was 23.80 years ( $SD = 7.07$ , range 18–55), and they were mostly well educated (40 of them had a university entrance diploma, 7 had a university degree, and 2 had an apprenticeship). They were Swiss ( $n = 38$ ), German ( $n = 9$ ), Swiss/German ( $n = 1$ ), or Slovenian ( $n = 1$ ). All were proficient or native German speakers.

### 2.1.2 Measures

#### 2.1.2.1 Cognitive probing of the self-defeating items

Participants completed the eight items of the SD scale in a randomized order (one item per page). Below each item, they were provided with a text box for the cognitive probing, which was designed according to the recommendations by

---

<sup>1</sup> The outcomes of the hypothesis tests were not altered by excluding these participants.



Behr et al. (2012). Participants were instructed to write down a short statement on “Why have you chosen this answer. Please provide a short explanation.”

#### 2.1.2.2 Content ratings

The content ratings of the cognitive probing statements were conducted by independent raters, who were not familiar with the study. They either rated the amount of self-directed humor or the state self-esteem of each of the 392 responses (always in randomized order). The amount of self-directed humor (defined as humor, amusement, jokes, and laughter about oneself) described in the response was rated on a scale of 1 “not at all,” 2 “somewhat,” and 3 “much.” The state self-esteem rating entailed whether the person exhibited a predominantly positive (+1), a predominantly negative (−1), or a neutral/balanced (0) evaluation of oneself in the response. Inter-rater reliability (or agreement between raters) was determined by the intra-class correlation based on a two-way random model with average measures ( $ICC_{2,2}$ ). The  $ICC_{2,2}$  is employed because (a) all raters coded all responses, (b) three-point Likert-type scales are used, and (c) because the average scores across the raters are of interest (as recommended by McGraw and Wong 1996). A value of 0 would indicate no agreement between the raters, while a value of 1 would indicate a perfect agreement between the raters. The values were sufficient for self-directed humor (3 raters,  $ICC_{2,2} = 0.90$ ) and state self-esteem (5 raters,  $ICC_{2,2} = 0.68$ ). The content ratings were averaged across all raters for each item response.

#### 2.1.2.3 Frequencies of emotion word

The responses were analyzed using a computer-assisted program for text analyses, the *Linguistic Inquiry and Word Count* (LIWC2001; Pennebaker et al. 2001), which can analyze texts based on its implemented dictionaries according to different categories. The categories of interest for the present studies are positive emotions (including words such as happy, good, love, and successful) and negative emotions (including words such as sad, hate, aggression, depressive). The analyses indicate the relative frequency (i.e., relative to the length of the text) of positive and negative emotion words.

### 2.1.3 Procedure

Participants from the preliminary study (Sample 2) who indicated an interest in participating in further online studies were invited via e-mail to participate in Study 1. They received an individual link with an ID to anonymously match their

responses from the preliminary study with Study 1. Participants completed the survey online ([www.unipark.info](http://www.unipark.info)) six to twelve weeks after the preliminary study. The content ratings were completed individually in the lab. Both participants and raters received course credit in psychology for their participation, and participants could also chose to receive general feedback on the study. Other variables were assessed that are not relevant for the present study.

### 2.1.4 Analyses

Gignac and Szodorai's (2016) empirical effect size guidelines for correlations in individual difference research were employed, categorizing effect sizes as  $|0.10|$  (relatively small),  $|0.20|$  (typical/medium), and  $|0.30|$  (relative large). The power was sufficient ( $\geq 0.80$ ) to detect small to medium effects ( $r \geq |0.13|$ ) across all 392 item responses and medium to large effects ( $r \geq |0.27|$ ) across the 105 affirmed item responses. Due to violations of the normal distribution, Spearman's rank correlations were computed. The word count in the single explanations ranged from 1–136 ( $M = 15.81$ ,  $SD = 13.07$ ). As the word counts were significantly correlated with the outcomes, partial correlations (controlling for the word counts) were computed.

## 2.2 Results and Discussion

### 2.2.1 Manipulation check

The two SD scales (from the preliminary study and from Study 1) were highly correlated ( $r_{[49]} = 0.92$ ,  $p < 0.001$ ). The scale means showed a significant, but negligible difference, with lower scores obtained in Study 1 ( $M = 22.08$ ,  $SD = 8.86$ ) compared to the preliminary study ( $M = 23.31$ ,  $SD = 8.66$ ;  $t_{[49]} = 2.45$ ,  $p = 0.018$ ,  $d = 0.14$ ). Additionally, the correlations between the SD scale and self-esteem and positive and negative affect (assessed in the preliminary study) were investigated. Replicating previous studies, the SD scale (Study 1) correlated negatively with the preliminary study measures of global self-esteem ( $r_{[49]} = -0.43$ ,  $p = 0.002$ ), positively with negative affect ( $r_{[49]} = 0.34$ ,  $p = 0.018$ ), and was not significantly correlated with positive affect ( $r_{[49]} = -0.22$ ,  $p = 0.125$ ). Overall, this shows that the cognitive probing paradigm of the SD items led to similar responses as the standard administration of the HSQ. Table 1 also shows the descriptive statistics of the variables assessed in Studies 1 and 2 (across all items and across the affirmed items).

Table 1: Means, Standard Deviations, Minimum and Maximum of the Measures in Study 1 and 2.

Measures	Study 1						Study 2			
	Responses to all items (N = 392)			Responses to affirmed items (n = 105)			Responses to all items (N = 104)			
	M	SD	Min	Max	M	SD	Min	Max	M	SD
Content ratings										
Self-directed humor	1.65	0.69	1.00	3.00	1.75	0.75	1.00	3.00	1.72	0.62
Self-esteem	0.20	0.48	-1.00	1.00	0.16	0.54	-1.00	1.00	0.21	0.50
Improving relationships									1.65	0.73
Word frequencies										
Positive emotions	5.77	9.41	0.00	100.00	7.11	10.26	0.00	66.67	4.45	4.69
Negative emotions	1.66	4.03	0.00	33.33	2.27	4.48	0.00	33.33	1.14	2.13
Frequencies of facial displays										
Positive emotions									0.20	0.43
Negative emotions									0.02	0.14

2.2.2 Item responses to different SD scores

The first set of hypotheses (1.1–1.3) deals with how item responses vary depending on the SD scores. It was expected that higher SD scores would be more maladaptive than lower ones. These hypotheses were tested with partial Spearman rank correlations, controlling for the word count (see Table 2).

As shown in Table 2, a small to medium positive correlation emerged between the SD item scores and the amount of self-directed humor described in the response, confirming Hypothesis 1.1. However, the correlations of the SD item scores with state self-esteem and the frequencies of negative emotion words were small and not significant, failing to support Hypotheses 1.2 and 1.3.

**Table 2:** Partial Spearman’s Rank Correlations (Controlling for Word Counts) of the Measures in Study 1 Across All Responses (N = 391–392, above the Diagonal) and Across the Affirmed Responses (5, 6, or 7 on a scale from 1–7; n = 105, below the Diagonal).

Measures	Content ratings		Word frequencies	
	Humor	Self-esteem	Positive emotions	Negative emotions
Self-defeating item score	0.12*	−0.05	0.09	0.05
Content ratings				
Self-directed humor	–	0.07	0.24***	−0.07
State self-esteem	0.38***	–	0.10	−0.15**
Word frequencies				
Positive emotions	0.34***	0.21*	–	−0.13**
Negative emotions	−0.12	−0.17	−0.11	–

Note. \*  $p < 0.05$ . \*\*  $p < 0.01$ . \*\*\*  $p < 0.001$ .

2.2.3 Connotations of the humor entailed in the SD items

The second set of hypotheses (1.4–1.6) deals with the connotations of the humor entailed in the SD items, which was expected to be either neutral or adaptive in terms of well-being. This was also tested with partial Spearman rank correlations (see Table 2). The amount of self-directed humor rated in each of the item responses correlated significantly and positively with the use of positive emotion words (medium effect), while the correlations with self-esteem and the use of negative emotion words were not significant. A similar pattern was found for the affirmed items ( $n = 105$ , or 26.8% of all responses), with the exception that self-directed humor was strongly positively associated with state self-esteem. This confirms Hypotheses 1.4–1.6, showing that the amount of self-directed humor entailed in the SD items was

either neutral or adaptive in terms of self-esteem and verbal emotion expressions (see Table 3 for examples of the responses and corresponding content ratings).

Overall, Study 1 showed that higher SD scores go along with humor, but not with a lower state self-esteem or with using negative emotion words. Additionally, the more self-directed humor was entailed in the item responses, the more positive emotion words were used and the higher self-esteem was (in affirmed items only). This is in line with the idea that there is an underlying negativity in the SD scale (as shown by the negative correlations of the SD scores with trait self-esteem and negative affect), but that the self-defeating humor itself is positive rather than negative. Thus, people with a lower self-esteem or a higher proneness to negative affect generally tend to score higher in the SD scale. Yet, if their responses include self-directed humor, they were rather

**Table 3:** Examples of the Responses and Content Ratings to the Item “I often go overboard in putting myself down when I am making jokes or trying to be funny” in Studies 1 and 2 (Translated from German).

SD score	Ratings			Response
	Humor	Self-esteem	IR	
Study 1: Online cognitive probing of typical scorers				
2	1.00	0.40		Never in a mean way towards myself.
3	2.00	0.40		Sometimes I don't notice that I go overboard, but not necessarily to make fun of someone. Just a cheerful mood.
7	1.00	0.60		I never pretend to be better than I actually am and prefer to understate than to come across as being arrogant.
Study 2: Thinking aloud of high scorers				
2	1.00	0.00	1.50	No, I can't think of anything right now, because I do not go overboard. Except for sometimes, especially if I do not feel so well or the like, that I additionally elaborate a bit more or something like that. So I choose "Mostly disagree." Because it is mostly not true that I go overboard.
5	1.00	0.00	1.00	Well, ok, what is "going overboard"? Well, I do go overboard a little but, somewhat, exaggerated, but not ... I don't know if you can call that "going overboard." Maybe ... "Partly agree" or "Neither agree or disagree." I chose "partly agree"

*Note.* SD = self-defeating, IR = improving relationships. Humor and IR rating from 1 “not at all” to 3 “much,” and self-esteem rating from –1 “negative self-evaluation” to 1 “positive self-evaluation.”

positively connoted, while they were less positively connoted when they did not entail humor. For example, responses to affirmed items rated low in humor and self-esteem contained explanations of not daring to say anything when being laughed at, not wanting to bother others or dragging others down, and becoming vulnerable by showing weaknesses or feelings, while responses rated high in humor and self-esteem pertained to liking to entertain others and making others laugh, not being bothered about one's mistakes or when others laugh at the person, accepting one's weaknesses, being more approachable by laughing at one's mishaps, and coping with humor.

While providing a detailed look into the interpretation of the SD scale and its items, Study 1 suffers from several shortcomings. First, although 392 item responses were analyzed, the distribution of the scores was similar to previous investigations (i.e., most items were disagreed to). Thus, the present findings might only hold for those with overall low to medium scores on the scale, while the high scorers (i.e., those who on average agree to the items) might show a more negative pattern. Second, the online cognitive probing led to rather short responses and thus a less than optimal data quality. Study 2 was thus designed to fill these gaps and to replicate and extend Study 1.

### 3 Study 2: Thinking Aloud of the Self-Defeating Humor Scale in High Scorers

Study 2 complements Study 1 (a) by investigating high scorers on the SD scale, (b) by employing the think-aloud procedure, and (c) by coding facial displays of emotion. The same hypotheses are set up as in Study 1, supplemented by hypotheses about the facial displays of emotion. Specifically, it is expected that the SD item scores correlate positively with the content rating of the amount of self-directed humor (Hypothesis 2.1) and negatively with the content rating of state self-esteem (Hypothesis 2.2). Regarding emotion expressions, it is expected that the SD item scores correlate positively with the use of negative emotion words (Hypothesis 2.3) and facial displays of negative emotions (Hypothesis 2.4). In terms of the humor entailed in the responses, it is expected that scores in the self-directed humor content rating are either uncorrelated or correlate positively with state self-esteem (Hypothesis 2.5), the use of positive emotion words (Hypothesis 2.6) and facial displays of positive emotion (Hypothesis 2.7), and that they are either uncorrelated or negatively correlated with the use of negative emotion words (Hypothesis 2.8) and facial displays of negative emotions (Hypothesis 2.9). Furthermore, no specific hypotheses are set up regarding the

degree to which improving interpersonal relationships is described in the item responses, because the definition of the self-defeating humor style includes the aim of enhancing one's relationships, while empirical findings supported negative relationships of the SD scale with traits such as antagonism and hostility.

## 3.1 Method

### 3.1.1 Sample

Participants from the preliminary study who indicated an interest in participating in a lab study and who had high scores in the SD scale (i.e., scores of  $\geq 33$  on a scale from 8–56) were invited to participate in Study 2. Of the 28 participants who had high scores in the preliminary study, 16 replicated their high score in Study 2. In the consistency check, three of them were excluded because their SD scores differed eight points or more from one another,<sup>2</sup> resulting in a final sample of 13 participants (SD scores ranging from 34–52). Five men and eight women participated, and their average age was 22.23 years ( $SD = 2.83$ , range 19–28 years). They were mostly well educated, with 11 having a university entrance diploma, one having a university degree, and one having a certificate of secondary education. Twelve were Swiss and one was German. All of them were native German speakers.

### 3.1.2 Measures

#### 3.1.2.1 Thinking aloud

Participants completed the eight SD items in a randomized order (each item was presented on a separate page). While answering the items, they continuously reported their thoughts (standard think-aloud procedure; see Beatty and Willis 2007). Specifically, participants were instructed to tell what they were thinking about while reading and answering the items. Before completing the SD items, they practiced thinking aloud with two unrelated items to make sure they understood the task. The think-aloud protocols were audio recorded and then transcribed.

#### 3.1.2.2 Content ratings and word frequencies

As in Study 1, independent raters judged either self-directed humor or state self-esteem contained in the 104 responses for each person and each item (always presented in a randomized order) on three-point scales. Inter-rater reliabilities

---

<sup>2</sup> The outcomes of the hypothesis tests were not altered by excluding these participants.

were sufficient for self-directed humor (2 raters,  $ICC_{2,2}=0.78$ ) and state self-esteem (3 raters,  $ICC_{2,2}=0.63$ ). Two additional raters judged whether the responses entailed an improvement of one's interpersonal relationships on a scale of 1 "not at all," 2 "somewhat," and 3 "much" ( $ICC_{2,2}=0.83$ ). Again, the relative frequencies of positive and negative emotion words contained in each response were analyzed using the LIWC2001.

### 3.1.2.3 Facial displays of emotion

Participants were unobtrusively filmed while thinking aloud. These videos were analyzed using an adapted version of the *Emotion Facial Action Coding System* (EMFACS-8; Ekman et al. 1994). The EMFACS is an observer-based scoring system for emotion-relevant facial behavior, based on coding specific action-unit (AU) combinations. Joy was coded if the cheeks were raised and the lip corners were pulled (AUs 6 + 12). Anger was coded if the brows were lowered (AU 4) and either the lips were tightened (AU 23) or pressed (AU 24), or if the lips were stretched and the jaw dropped (AUs 20 + 26, with AU 26 being stronger than AU 20). Sadness was scored when the inner brows were raised and the lip corners were depressed (AUs 1 + 15). Finally, fear was scored if either the inner and outer brows were raised and the brows were lowered (AUs 1 + 2 + 4) or if the upper lid was raised (AU 5), co-occurring with stretched lips and lips parting (AUs 20 + 25) or with stretched lips and a jaw drop (AUs 20 + 26, with AU 20 being stronger than AU 26). The AUs were also scored on a five-point intensity scale (from A = slight to E = maximum). To code a relevant AU combination, the involved AUs needed to have a joint apex (i.e., their maximum intensities needed to occur simultaneously) and their intensities needed to be similar (i.e., not more than two intensity levels apart).

After coding all relevant AUs and their combinations, full emotion displays were summed up separately for displays of positive emotions (i.e., joy) and negative emotions (i.e., anger, fear, and sadness) for each item response. Two certified FACS coders independently conducted the EMFACS codings and control codings. Inter-coder reliability was compared across eight SD items, yielding sufficient Cohen's kappas for the EMFACS events of positive (0.68) and negative emotions (0.68).

### 3.1.3 Procedure

Participants first provided their written informed consent to participate and to have their utterances audio recorded. After the study, participants were debriefed about the nature of the study and the video recordings, and they provided additional written consent regarding the usage of their audio and video recordings.



Participants were alone when completing the tasks, and the experimenter was connected via intercom to answer any questions. As in Study 1, raters completed the content ratings individually in the lab. Participants received either course credit in psychology or a remuneration of 30 Swiss franks, and raters received course credit in psychology. The ethics committee of the Faculty of Arts at the University of Zurich approved the study. Other variables were collected that are not relevant for the present study.

### 3.1.4 Analyses

The same effect size guidelines as in Study 1 were employed. Power was sufficient ( $\geq 0.80$ ) to detect medium to large effects ( $r \geq |0.27|$ ). The word count of the explanations of the SD items ranged from 6–191 words ( $M = 53.17$ ,  $SD = 38.86$ ). As in Study 1, the word counts correlated with the other measures, and hence partial correlations (Spearman's rank correlations) were computed.

## 3.2 Results and Discussion

### 3.2.1 Manipulation check

The SD scale from the preliminary study and Study 2 were highly correlated ( $r_{[13]} = 0.79$ ,  $p = 0.001$ ) and their means were not significantly different from one another (preliminary study:  $M = 38.93$ ,  $SD = 5.43$ , Study 2:  $M = 38.77$ ,  $SD = 5.43$ ;  $p = 0.841$ ). As expected, in comparison to the sample from Study 1, the SD score ( $z_{[13]} = 4.82$ ,  $p < 0.001$ ,  $d = 1.55$ ) was significantly higher in Study 2. Comparing participants' well-being scores from the preliminary study, positive affect was lower in Study 2 ( $M = 3.18$ ,  $SD = 1.12$ ) than in Study 1 ( $M = 3.73$ ,  $SD = 0.79$ ;  $z_{[13]} = 2.07$ ,  $p = 0.038$ ,  $d = 0.55$ ), and negative affect was higher in Study 2 ( $M = 3.36$ ,  $SD = 1.02$ ) than in Study 1 ( $M = 2.77$ ,  $SD = 0.70$ ;  $z_{[13]} = 2.14$ ,  $p = 0.032$ ,  $d = 0.57$ ), while no significant difference was found for self-esteem (Study 1:  $M = 3.03$ ,  $SD = 0.71$ ; Study 2:  $M = 2.92$ ,  $SD = 0.80$ ;  $z_{[13]} = 0.40$ ,  $p = 0.689$ ). Thus, the high scorers in Study 2 indeed scored higher in the SD scale and negative affect and lower in positive affect than the standard sample in Study 1.

### 3.2.2 Item responses to different SD scores

As in Study 1, the first set of hypotheses (2.1–2.4) deals with how item responses vary depending on the SD scores, expecting that higher SD scores would be

more maladaptive than lower ones. These hypotheses were tested with partial Spearman rank correlations, controlling for the word count (see Table 4).

As shown in Table 4, the scores in the SD items showed medium to large positive correlations with the state self-esteem and improving relationships content ratings and with the frequencies of facial displays of positive emotions. These findings fail to support Hypotheses 2.1–2.4.

**Table 4:** Partial Spearman’s Rank Correlations (Controlling for Word Counts) of the Measures in Study 2 Across All Item Responses.

Scales	Content ratings			Word frequencies		Facial displays	
	Humor	Self-esteem	IR	PE	NE	PE	NE
Self-defeating item score	0.08	0.24*	0.34***	0.09	0.01	0.28*	−0.12
Content ratings							
Self-directed humor		0.32**	0.20*	0.23**	−0.02	0.13	−0.09
Self-esteem			0.35***	0.21*	−0.05	0.04	−0.05
Improving relationships (IR)				0.12	0.12	0.00	−0.10
Word frequencies							
Positive emotions (PE)					−0.03	0.12	−0.19
Negative emotions (NE)						−0.11	0.16
Frequencies of facial displays							
Positive emotions (PE =							−0.05

Note. N = 104. \*p < 0.05. \*\*p < 0.01. \*\*\*p < 0.001.

3.2.3 Connotations of the humor entailed in the SD items

As in Study 1, the second set of hypotheses (2.5–2.9) deals with the connotations of the humor entailed in the SD items, which was expected to be either neutral or adaptive in terms of well-being. This was also tested with partial Spearman rank correlations (see Table 4). The amount of self-directed humor entailed in the responses to the SD items was positively related to self-esteem (large effects) and to improving relationships (medium effect) as well as to the frequency of using positive emotion words (medium effect).<sup>3</sup> This confirms Hypotheses 2.5–2.9.

3 Similar correlations were obtained across the affirmed items, which is expected given that most items were agreed to (n = 72, or 69.2% of all responses).

Thus, self-directed humor as entailed in the SD scale was associated with increased positive outcomes (a higher state self-esteem and improving relationships, and using more positive emotion words), but was unrelated to negative outcomes (i.e., expressions of negative emotions). This is in line with the recent studies that supported the notion that the humor entailed in the SD scale was either neutral or positive in terms of psychological well-being. Study 2 provides a replication for high scorers on the SD scale and extends the findings to relevant social aspects (i.e., facial emotion expressions and interpersonal relationships).

## 4 General discussion

The present studies employed cognitive interviewing techniques to investigate the interplay of the SD scale and the humor entailed in it with self-esteem and emotions. In Study 1, higher SD scores contained more self-directed humor than lower scores. In Study 2, higher SD scores contained a higher state self-esteem, more improving relationships, and they were accompanied by more facial displays of positive emotion than lower scores. Employing two methods of cognitive interviewing techniques, covering low, medium and high scorers on the SD scale, and employing multiple assessment methods (self-reports, content ratings, and two behavior-based measures) supports the generalizability of these findings. Thus, Martin et al.'s (2003) conceptualization of the self-defeating humor style as a combination of humor at the expense of self to enhance one's relationships could only be partially supported. Specifically, humor and improving relationships was supported, while lower state self-esteem (i.e., expense and detriment of oneself) was not supported.

This could be because people differed in how they interpreted the SD items in terms of humor and self-esteem. Specifically, the more self-directed humor they described, the higher the state self-esteem was rated and the more positive emotion words were employed (while this effect was not significant for the facial displays of emotions). In other words, instead of suggesting that self-defeating humor contributes to lower psychological well-being (as was done in previous studies), it rather might to be the case that this humor is employed to enhance the psychological well-being of people who have a lower self-esteem or a proneness to negative affect. In this sense, self-defeating humor might serve a self-enhancing or coping function in dealing with mistakes, weaknesses, and problems. For example, participants described self-defeating humor as a good coping mechanism to deal with depression and anxiety, and described that it helped to admit, to accept, and to see the funny side of one's shortcomings and one's imperfections.

Although not a focus of the present studies, the content ratings in Study 2 also showed that the more self-directed humor was described in the item responses, the more the responses also entailed improving interpersonal relationships. Thus, self-directed humor might not only have positive effects for the person showing it, but it might also help in fostering their relationships. For example, participants mentioned that the self-directed humor helped them to cheer others up who were anxious or in a bad mood, to deal with problems, to lighten up tense and stressful situations with others, to become more approachable (i.e., “breaking the ice”) and to strengthen social bonds. This is also in line with Janes and Olson’s (2015: 274–275) notion that “self-deprecating humor tends to be more lubricating than abrasive, because it does not make anyone feel defensive.” While previous studies found the SD scale to be associated with maladaptive interpersonal outcomes (such as hostility and antagonism), this suggests that self-defeating humor could actually serve a positive social function, in line with the conceptualization of this humor style.

## 4.1 Implications for Future Research and Applications

Several important implications can be derived from the present findings. First, they provide further and more direct support for the idea that the SD scale consists of a positive self-directed humor exhibited by people with lower self-esteem or a proneness to negative affect. Thus, previous findings of the negativity of the SD scale cannot be interpreted in terms of the (rather positive) humor that is entailed in it, and no causal role can be attributed to the humor in terms of well-being. Hence, both the conceptualization and measurement of the self-defeating humor style need amendments before they can be meaningfully employed in both research and applications. Otherwise, potentially detrimental conclusions might be drawn, such as decreasing self-defeating humor instead of fostering it to enhance well-being. Second, future research could directly compare different forms of self-directed humor, for example, the self-defeating humor style, laughing at oneself, and self-deprecating humor, to elucidate their overlaps and differences. In these analyses, the “shotgun wedding” between the negative self-evaluation and the self-directed humor entailed in the self-defeating humor style would need to be undone to clarify both positive and negative aspects of this construct.

Third, establishing self-directed humor as mostly positive in terms of psychosocial well-being presents an opportunity to include it in humor interventions or trainings. McGhee’s (2010) humor habits program already entails laughing at oneself as one of seven humor habits. While the program was shown to effectively enhance psychological well-being and to decrease maladaptive outcomes (for an

overview, see Ruch and McGhee 2014), studies on the specific effects of self-directed humor and laughing at oneself are still needed.

## 4.2 Limitations

Several limitations must be noted in the present set of studies. First, the samples employed were rather small and not representative for adults in general (e.g., mostly females in Study 1 and mostly well-educated participants in Study 2), and thus replications with larger and more representative samples are necessary to be able to generalize the present findings. Also, replications of the findings in other languages and cultures would be desirable. Second, more fine-grained analyses of the responses and the facial displays could be conducted in future studies. For example, self-conscious emotions could be incorporated (such as pride and shame; see Tracy et al. 2007) or the FACS (Ekman et al. 2002) could be used to detect micro expressions (i.e., facial displays of emotions that only last 15–25 msec.). Third, eliciting spontaneous self-defeating humor in an experimentally controlled design and comparing it with other kinds of (self-directed) humor would provide information on the causal relationships of humor and well-being.

## 4.3 Conclusions

The present set of employed cognitive interviewing techniques of the SD scale to yield information on the interplay between self-defeating humor, state self-esteem, and emotion expressions. Both studies found a contrast between the maladaptive connotation of the SD scale and the humor entailed in it. Self-defeating humor was found to be neutral or positive in terms of self-esteem, emotions, and also interpersonal relationships. Thus, engaging in self-directed humor by making fun of oneself, by being amused about one's own shortcomings and mishaps, and by laughing at oneself, even if done excessively, seems to serve a coping function and should thus be better encouraged than discouraged. Overall, this opens new perspectives for research and applications of the potential positive effects of self-directed humor for both oneself and for others.

**Acknowledgements:** We would like to thank Jenny Hofmann for her helpful comments on earlier versions of the manuscript, and Yoel Everett for the control codings of the facial displays of emotion. Furthermore, we would like to thank the “Stiftung Humor und Gesundheit” Switzerland that financially supported parts of the data collection (grant no. 160'229 awarded to SH).

**Funding:** “Stiftung Humor und Gesundheit” Switzerland, Grant Number: 160’229

## References

- Beatty, Paul C & Gordon B Willis. 2007. Research synthesis: The practice of cognitive interviewing. *Public Opinion Quarterly* 71(2). 287–311.
- Behr, Dorothée, Lars Kaczmarek, Wolfgang Bandilla & Michael Braun. 2012. Asking probing questions in web surveys: Which factors have an impact on the quality of responses? *Social Science Computer Review* 30(4). 487–498.
- Bosson, Jennifer K., William B Swann Jr & James W. Pennebaker. 2000. Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited? *Journal of Personality and Social Psychology* 79(4). 631–643.
- Diener, Ed, Derrick Wirtz, William Tov, Chu Kim-Prieto, Dong-won Choi, Shigehiro Oishi & Robert Biswas-Diener. 2010. New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social Indicators Research* 97(2). 143–156.
- Ekman, Paul, Wallace V Friesen & Joseph C Hager. 2002. *Facial Action Coding System: The manual*. Salt Lake City: Research Nexus.
- Ekman, Paul, William Irwin & Erika L Rosenberg. 1994. *EMFACS-8: Coders Instructions*. Unpublished manuscript, University of California, San Diego.
- Ford, Thomas E., Shaun K Lappi, Emma C O'Connor & Noely C Banos. 2017. Manipulating humor styles: Engaging in self-enhancing humor reduces state anxiety. *Humor: International Journal of Humor Research* 30(2). 169–192.
- Gignac, Gilles E & Eva T Szodorai. 2016. Effect size guidelines for individual differences researchers. *Personality and Individual Differences* 102(11). 74–78.
- Heintz, Sonja. 2017. Putting a spotlight on daily humor behaviors: Dimensionality and relationships with personality, subjective well-being, and humor styles. *Personality and Individual Differences* 104(1). 407–412.
- Janes, Leslie & James Olson. 2015. Humor as an abrasive or a lubricant in social situations: Martineau revisited. *Humor: International Journal of Humor Research* 28(2). 271–288.
- Kuiper, Nicholas A. (ed.). 2014. Humor, well-being and health. [Special issue]. *Europe's Journal of Psychology* 10(3).
- Maiolino, Nadia B & Nicholas A Kuiper. 2014. Integrating humor and positive psychology approaches to psychological well-being. *Europe's Journal of Psychology* 10(3). 557–570.
- Martin, Rod A., Patricia Puhlik-Doris, Gwen Larsen, Jeanette Gray & Kelly Weir. 2003. Individual differences in uses of humor and their relation to psychological well-being: Development of the Humor Styles Questionnaire. *Journal of Research in Personality* 37(1). 48–75.
- Matsumoto, David, Dacher Keltner, Michelle N. Shiota, Maureen O'Sullivan & Mark Frank. 2010. Facial expression of emotion. In Jeannette M Michael Lewis, Haviland Jones & Lisa F Barrett (eds.), *Handbook of emotions* (3rd ed.), 211–234. New York: Guilford Press.
- McGhee, Paul E. 2010. *Humor as survival training for a stressed-out world: The 7 humor habits program*. Bloomington: AuthorHouse.
- McGraw, Kenneth O & Seok P. Wong (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods* 1(1). 30–46.

- Pennebaker, James W., Martha E Francis & Roger J Booth. 2001. *Linguistic Inquiry and Word Count – LIWC2001*. Mahwah: Erlbaum.
- Rnic, Katerina, David J. A. Dozois & Rod A. Martin. 2016. Cognitive distortions, humor styles, and depression. *Europe's Journal of Psychology* 12(3). 348–362
- Rosenberg, Morris. 1965. *Society and the adolescent self-image*. Princeton: Princeton University Press.
- Rosenberg, Morris, Carmi Schooler, Carrie Schoenbach & Florence Rosenberg. 1995. Global self-esteem and specific self-esteem: Different concepts, different outcomes. *American Sociological Review* 60(1). 141–156.
- Ruch, Willibald & Sonja Heintz. 2013. Humour styles, personality, and psychological well-being: What's humour got to do with it? *European Journal of Humour Research* 1(4). 1–24.
- Ruch, Willibald & Sonja Heintz. 2017. Experimentally manipulating items informs on the (limited) construct and criterion validity of the Humor Styles Questionnaire. *Frontiers in Psychology: Personality and Social Psychology* 8(616).
- Ruch, Willibald & Paul E McGhee. 2014. Humor intervention programs. In Acacia C Parks & Stephen M Schueller (eds.), *The Wiley Blackwell handbook of positive psychological interventions*, 179–193. Chichester: John Wiley & Sons.
- Tracy, Jessica L., Richard W Robins & June P Tangney (eds.). 2007. *The self-conscious emotions: Theory and research*. New York: Guilford Press.
- Zeigler-Hill, Virgil, Gillian A McCabe & Jennifer K Vrabell. 2016. The dark side of humor: DSM-5 pathological personality traits and humor styles. *Europe's Journal of Psychology* 12(3). 363–376.

## Bionotes

### Sonja Heintz

Sonja Heintz is a postdoctoral researcher at the Department of Psychology at the University of Zurich. Her main research interests are individual differences in humor (humor styles, humor behaviors, and humor appreciation/production), the measurement of humor, and positive psychology (e.g., character strengths and well-being).

### Willibald Ruch

Willibald Ruch is a Full Professor of Psychology at the University of Zurich. His research interests are in the field of humor and laughter, cheerfulness and smiling. His recent work, together with his research team at the University of Zurich, includes humor from a positive psychology perspective, the effectiveness of humor training programs and clown interventions, the fear of being laughed at (i.e., gelotophobia), and the measurement of humor.